DOCUMENT RESUME

ED 084 514 CS 000 847

AUTHOR Anderson, William W.
TITLE Evaluation of College Reading and Study Skills Programs; Problems and Approaches.
PUB DATE Nov 73
NOTE 11p.; Paper presented at the Annual Meeting of the College Reading Assn. (17th, Silver Spring, Md., Nov. 1 through 3, 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Affective Behavior; *College Students; Grade Point Average; Grade Prediction; Higher Education; Measurement Techniques; Reading Development; *Reading Improvement; *Reading Programs; *Reading Research; Reading Skills; *Reading Tests; Test Interpretation

ABSTRACT

The purpose of this paper was to bring together some of the most outstanding and significant research in recent years relating to the problems of evaluating college reading and study skills programs and the corresponding prospects and implications for improvement. The sections of the paper include: "Criterion Measures of Program Success," which is a discussion of the use of grade-point average (GPA) as a criterion measure for reading program success; "Measurement Tools," which is a look at the variety of standardized tests available for college reading improvement programs (primarily their inadequancies); "Statistical Design of Evaluative Studies," which is a discussion of some of the statistical problems in program evaluative studies and other design problems; "Affective Variables in Evaluation," which is a description of the predictive value of GPA in relation to affective measures and some of the problems inherent in affective measurement; and "Summary and Conclusions," which is a listing of four factors--standardized test limitation; new developments in statistical design techniques; disappearance of canned, ultra-behavioristic commercial programs; and a growing research bank in college reading--that may ultimately improve reading improvement programs. (WR)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED 084514

EVALUATION OF COLLEGE READING AND STUDY SKILLS PROGRAMS

PROBLEMS AND APPROACHES

William W. Anderson
Shippensburg State College
Shippensburg, Pennsylvania

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
William W. Anderson

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER

Paper presented at the
College Reading Association
Seventeenth Annual Meeting,
Silver Spring, Maryland,
November 2, 1973

CS 000 847

FILMED FROM BEST AVAILABLE COPY

INTRODUCTION

The past five years in college reading and study skills instruction have been characterized by rapid growth and renewed interest. There has been a proliferation of new programs and corresponding expansion of existing ones. The obvious trend toward declining undergraduate enrolments has resulted in lower admissions standards or, in many cases, "open-admissions" policies. This trend, in combination with the rapid expansion and popularity of the community college, has produced an almost overwhelming influx of underprepared students. Without adequate programs to provide assistance in reading and study skills a large percentage of these students are doomed to frustration, disillusionment and failure. In an attempt to meet these needs, many institutions have hastily developed token, "finger-in-the-dike" type programs that seem to do little more than to provide a placebo for their clientel. Happily, however, such developments have precipitated renewed research attention focused upon the evaluation and improvement of college reading and study skills programs. Despite the gratifying progress that has been achieved in recent years, a review of the literature makes it clear that the state of the art in college reading program-evaluation is one of becoming rather than being.

PURPOSE

The purpose of this paper is to bring together some of the most outstanding and significant research in recent years relating to the problems of evaluating college reading and study skills programs and the corresponding prospects and implications for improvement.

The research literature seems to reflect the need (1) for use of more nearly adequate and appropriate criterion measures of success, (2) for improved measurement tools, (3) for more sophisticated and appropriate statistical design in program evaluation studies, and (4) for greater attention to affective variables as predictors of success. Each of these factors will be discussed briefly as it relates to college reading and study skills program evaluation.

CRITERION MEASURES OF PROGRAM SUCCESS

Though short-term objectives should be varied according to individual needs and assorted local variables, the most obvious, and most defensible long-range objective is the improvement of grade-point average (GPA). The evidence suggests, however, that systematic use of GPA as the major criterion measure is rare (Maxwell, 1971). For example, a survey of seventeen California Community Colleges by Berg and Axtel (1968) revealed that only one had a systematic program for assessment of GPA. Entwistle and Entwistle (1960) summarize fourteen studies that used GPA as a criterion measure. They report a modal GPA gain of .4 to .5. Moreover, they indicate that improvement was almost always maintained. The Entwistles (1960) emphasize that the only adequate criterion measure is GPA improvement in comparison with a control group.

Those studies that do use subsequent GPA as a criterion measure for reading program success report mixed results. A review of those investigations that appear to be most carefully controlled indicates rather consistent GPA improvement in highly verbal subjects but there is little reliable indication of improved GPA performance in quantitative courses (Willey & Thompson, 1956; Gunderson, 1961; Wright, 1962). Among those studies reviewed that looked at verbal and quantitative subjects separately, there were no reports found of

quantitative GPA improvement. Tillman (1973) reviewed 31 studies of four-year college reading improvement programs and grades which were reported in the literature between 1945 and 1971. Among these he found 23 that reported statistically significant gain in GPA to be associated with participation in a reading program. He further notes that grade gains reported "tended to be practically small" (1973, p. 100).

Maxwell (1971) suggests that overall GPA may not be an appropriate criterion measure and that perhaps a GPA excluding science and mathematics subjects would be a more valid measure of program effectiveness. Both Maxwell (1969, 1971) and Tillman (1973) question the wisdom of exclusive reliance upon GPA as a criterion measure and suggest the use of such additional criterion variables as attendance, informal feedback, attrition from enrolment, persistence of gains, etc.

MEASUREMENT TOOLS

So much attention has been recently directed toward the inadequacies of standardized testing that it almost seems trite to elaborate on this point. Yet it will perhaps be of some value to beiefly examine the rather glaring weaknesses of those instruments most commonly used in college reading improvement programs. Tillman (1973) reports that he found the Iowa Silent Reading Test, the Nelson-Denny, and the Cooperative Reading Test to be among the three most frequently used.

The Iowa Silent Reading Test is now nearing completion of a badly-needed, thorough revision. The older edition used in so many of the earlier studies suffered from a variety of weaknesses. Among those that seem most apparent are the awkward and unappealling format, the remote relationship of several of the subtests to actual reading situations, and cutdated norms. Recognition of these weaknesses is suggested in the research literature

reporting a decreasing frequency of usage of this test in more recent years (Tillman, 1973). Even given all of the weaknesses of more modern tests, when viewed in the perspective of some of the improvements that they offer, the Iowa Silent seems to be almost anachronistic.

The Nelson-Denny Reading Test also suffers from a host of inadequacies. A review by Farr (1972) cites insufficient reliability data, inadequate validity evidence, an excessively broad range of measurement span (i.e. eight grade levels), content biased to favor those with literary interests, and the confounding of scores purporting to represent vocabulary and comprehension power by severe time limitations. Moreover, his review even implies inaccuracy and exaggeration of the reliability data. Previous studies of minimal time (i.e. 4 minutes) required to measure reading rate with acceptable validity and reliability (Humphry, 1957) cast doubt upon the value of the one-minute rate test on the Nelson-Denny.

The Cooperative Reading Test appears to be the most acceptable of the three in common use at the college level. The reviewer (Lahey, 1972) indicates several notable strengths including rather complete reliability, validity and norming data and the reporting of scores in percentile bands ... This test utilizes the levels of comprehension concept and is unique as such. The most glaring weakness seems to be the use of isolated words for measuring vocabulary development.

It is interesting to note that though Davis (1961) strongly recommends that only tests scored with a correction for chance success be employed to estimate improvement resulting from learning, none of the three tests mentioned above are so scored. Davis also suggests that test directions are inadequate without some explanation of how scoring is to be done. None of the tests most frequently used in college reading improvement programs adequately qualifies on this suggested criteria for test selection.

Another glaring weakness of standardized tests for college reading improvement programs is their failure to measure many of the primary skill goals of such programs. Skills such as purpose, setting, flexibility and retention are often ignored by tests (Maxwell, 1971).

STATISTICAL DESIGN OF EVALUATIVE STUDIES

The problem so apparent in many reading research studies of inadequate control for Hawthorne Effect, regression to the mean and other uncontrolled, spurious variables is especially notable in this area. Only a small percentage of the studies reviewed controlled for Hawthorne Effect. McDonald (1964) estimated that more than eighty (80) per cent of the studies dealing with evaluation of progress in reading programs suffer from serious contamination due to the Hawthorne and placebo effects. In a review of thirty-one (31) studies purporting to measure the relationship between reading training and college success Wright (1962) found only eleven (11) with comparable control groups. Many of those studies having control groups failed to control for motivation due to the voluntary nature of experimental programs. A slightly more optimistic picture is seen by Tillman (1973) who notes with interest that all of the thirty-one (31) studies that he reviewed either used a no-treatment group or attempted statistical control. Most of the studies that he reviewed accomplish this control by matching subjects with only a few using covariance techniques.

Perhaps the most difficult and persistent statistical problem in program-evaluative studies has been regression toward the mean. All of the studies reviewed reported reading improvement and GPA improvement in terms of "crude gain". Tracy and Rankin (1967) suggest that the "residual gain" technique developed by Manning and DuBois (1962) be used to remove the influence of regression effects on measurement of improvement. This technique equates the

subjects statistically on the basis of pretraining measurement. It controls for initial status, regression of errors and other problems of "crude gain."

Another basic design problem pointed out by Maxwell (1971) is the inappropriate nature of statistics that assume a normal distribution for the typically skewed population of less able readers.

AFFECTIVE VARIABLES IN EVALUATION

If the primary reading program objective is to be the improvement of GPA, there is evidence to suggest that affectively focused programs may be more effective than those with cognitive emphasis. A longitudinal study of predictors of academic success among high-risk and regular college freshmen at Shippensburg State College found that an appreciably greater amount of variance in subsequent GPA performance can be accounted for through personality measures than through ability measures. A comparison study of the GPA predictive value of personality and ability measures with three groups of college freshmen using step-wise multiple linear regression techniques was conducted. Ability measures (i.e. SAT scores and class rank) were found to account for between seven (7) and eighteen (18) per cent of the variance in subsequent GPA performance whereas a personality measure, the Adjective check List, was able to account for sixteen (16) to thirty-four (34) per cent of the variance among the groups (Burnett, 1973). Studies by several other researchers support this finding (Miller & O'Connor, 1969). Though no studies were reviewed that compared the GPA predictive value of reading achievement tests with affective measures, it would seem to follow that they would be no more predictive of college success than SAT scores.

Though the evidence from which to draw such conclusions is yet rather meager, that which is available would seem to suggest that measures of the affective quality of reading and study skills improvement programs may well be

more appropriate than strictly cognitive concerns. Such measures might include the Adjective Check List (Gough and Heilbrun, 1965), Brown-Holtzman Survey of Study Habits and Attitudes, the Estes Scale for Measuring Reader Attitude (Estes, 1971), etc. Consideration of greater focus upon affective dimensions of reading improvement programs seems to be particularly appropriate in view of the number of studies that are reasonably well-controlled parcelling out I.Q. differences, etc., that indicate no significant difference in GPA in relation to reading achievement test gains.

There are, of course, numerous problems inherent in affective measurement also. Much of the research identifying the affective characteristics of high-risk students is primarily based upon correlation techniques. The popularly used Brown-Holtzman Survey of Study Habits and Attitudes seems to yield spuriously low scores due to rapidly changed social attitudes among adolescents since the gathering of normative data in the early sixties.

SUMMARY AND CONCLUSIONS

Though many reading professionals find the continuing gap between research and practice to be exasperating, though many feel pressured by deans to produce data concerning their programs that is shallow yet impressive to the layman, though the heavily vested interests of producers of assorted gadegetry and simplistically conceived tests still strongly influence college reading improvement programs, there remain many reasons for optimism. These include, but are not limited to:

(1) Growing awareness of standardized achievement test limitations and correspondingly greater attention to various alternative evaluative techniques.

(2) Rapid developments in statistical design techniques such as "residual gain", Bayesian concepts, etc., that provide the tools for rapid growth of knowledge in college reading improvement instruction.

(3) Slow disappearance of canned, ultra-behavioristic, commercial programs that dominated the field in years gone by in favor of more cognitive and humanistically based programs.

(4) A rapidly growing research bank in college reading represented by such notable publishing efforts as the National Reading Conference Yearbook summaries by Bliesmer.

REFERENCES

Berg, E. H. and Axtel, D. Programs for disadvantaged students in the California community colleges. Peralta Junior College District, 1968.

Burnett, B. Personality and ability variables as predictors of academic success for both high-risk and regularly admitted college students. (Master's Thesis, Shippensburg State College, Shippensburg, PA.) 1973.

Davis, F. B. The assessment of change. In Emery P. Bliesmer and Albert J. Kingston (Eds.), Phases of college and other adult reading programs (Tenth yearbook of the National Reading Conference), Milwaukee, Wis., The National Reading Conference, 1961, 86-95.

Entwistle, C. and Entwistle, R. Study-skills courses in medical schools. Journal of Medical Education, 1960, 35 (9).

Estes, T. H. A scale to measure attitudes toward reading. Journal of Reading, 1971, 15, 135-138.

Farr, R. Review of Nelson-Denny Reading Test. In Wm. Blanton, Roger Farr and J. Jaap Tuinman (eds.), Reading tests for the secondary grades: a review and evaluation. Newark, Del., International Reading Association, 1972, 31-34.

Gough, H. G. and Heilbrun, A. B., Jr. The Adjective Check List Manual. Palo Alto: Consulting Psychologist Press, 1965.

Gunderson, D. V. The influence of college reading instruction upon academic achievement. (Doctoral dissertation, University of Minnesota) Ann Arbor, Mich.: University Microfilms, 1961, No. 60-5610.

Humphrey, K.H. An investigation of the amount-limit and time-limit methods of measuring rate of reading. Journal of Developmental Reading, 1957, 1, 41-54.

Lahey, V.M. Review of cooperative English Tests -- Reading section. In Wm. Blanton, Roger Farr and J. Jaap Tuinman (eds.), Reading Tests for the secondary grades: a review and evaluation. Newark, Del., International Reading Association, 1972, 12-15.

Manning, W.H. and DuBois, P.H. correlational methods in research on human learning. Perceptual and Motor Skills, 1962, 15, 287-321.

Maxwell, M.J. Predicting persistence in a voluntary college reading program. (Paper read at the International Reading Association Convention) 1969.

Maxwell, M.J. Evaluating college reading and study skills programs. Journal of Reading, 1971, 15, 214-221.

Maxwell, M.J. Evaluating an individualized college reading service: alternative criteria and statistical models to the test-retest paradigm. Journal of Reading Behavior, 1972, 4 (3), 1-8.

McDonald, A.S. Some pit-falls in evaluating progress in reading instruction. In Roger Farr (ed.), Measurement and evaluation of reading. New York: Harcourt, Brace, Javanovich; 1970, 358-365.

Miller, D.N. and O'Connor, P. Achiever personality and academic success among disadvantaged college students. Journal of Social Issues, 1969, 25, 103-116.

Tillman, C.E. Four-year college reading improvement programs and grades: an annotated review, 1945-1971, Journal of Reading Behavior, 1973, 5, 100-109.

Tracy, R.J. and Rankin, E.F., Jr. Methods of computing and evaluating residual gain scores in the reading program. Journal of Reading, 1967, 10, 363-371.

Willey, D.S. and Thompson, C.W. Effective reading and grade-point improvement with college freshmen. School and Society, 1956, 83, 134-135.

Wright, E.S. The effect of reading training on college achievement. In Alton L. Raygor (Ed.) College and adult reading, First Annual Yearbook of the North Central Reading Association. Minneapolis, Minn., The North Central Reading Association, 1962, 32-50.